



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2015

Interruptibility of software developers and its prediction using psycho-physiological sensors

Züger, Manuela ; Fritz, Thomas

Abstract: Interruptions of knowledge workers are common and can cause a high cost if they happen at inopportune moments. With recent advances in psycho-physiological sensors and their link to cognitive and emotional states, we are interested whether such sensors might be used to measure interruptibility of a knowledge worker. In a lab and a field study with a total of twenty software developers, we examined the use of psycho-physiological sensors in a real-world context. The results show that a Naive Bayes classifier based on psychophysiological features can be used to automatically assess states of a knowledge worker's interruptibility with high accuracy in the lab as well as in the field. Our results demonstrate the potential of these sensors to avoid expensive interruptions in a real-world context. Based on brief interviews, we further discuss the usage of such an interruptibility measure and interruption support for software developers.

DOI: <https://doi.org/10.1145/2702123.2702593>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-110157>

Conference or Workshop Item

Accepted Version

Originally published at:

Züger, Manuela; Fritz, Thomas (2015). Interruptibility of software developers and its prediction using psycho-physiological sensors. In: CHI 2015, Seoul, South Korea, 18 April 2015 - 23 April 2015, ACM.

DOI: <https://doi.org/10.1145/2702123.2702593>

Interruptibility of Software Developers and its Prediction Using Psycho-Physiological Sensors

Manuela Züger and Thomas Fritz

Department of Informatics
University of Zurich, Switzerland
{zueger, fritz}@ifi.uzh.ch

ABSTRACT

Interruptions of knowledge workers are common and can cause a high cost if they happen at inopportune moments. With recent advances in psycho-physiological sensors and their link to cognitive and emotional states, we are interested whether such sensors might be used to measure interruptibility of a knowledge worker. In a lab and a field study with a total of twenty software developers, we examined the use of psycho-physiological sensors in a real-world context. The results show that a Naïve Bayes classifier based on psycho-physiological features can be used to automatically assess states of a knowledge worker's interruptibility with high accuracy in the lab as well as in the field. Our results demonstrate the potential of these sensors to avoid expensive interruptions in a real-world context. Based on brief interviews, we further discuss the usage of such an interruptibility measure and interruption support for software developers.

Author Keywords

psycho-physiological, lab/field study, interruptibility

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

In office work environments, interruptions by co-workers, emails or instant messages are common. While some of these interruptions are desired, others might incur a high cost, including long resumption lags, slower performance, negative emotions and an increase in errors due to the interruption happening at inopportune moments [6, 16]. When a co-worker is asked to assess a colleague's interruptibility into five states—from highly interruptible to highly non-interruptible—the assessment is difficult and only slightly better than chance as Fogarty et al. found in their study [17]. In addition, these assessments are generally based on cues of the colleague's social and task engagement, such as an open door, the colleague talking to someone or the use of the computer keyboard. However, especially in today's globally distributed work environments, this context information is often not available.

To avoid the high costs that interruptions can cause on knowledge workers, researchers have looked at automatically identifying good and bad moments for interruptions and computing a measure for a worker's interruptibility. Such an automatic interruptibility measure can then be used to better coordinate interruptions by, for instance, providing visual cues or postponing them to a more suitable moment [37]. Prior work examined the use of context-aware sensors to gather information, such as audio and video streams, keyboard and mouse interaction, or task characteristics (e.g., [17, 18, 27]).

With recent advances in psycho-physiological (aka. biometric) sensor technology, researchers have also started to investigate their use to assess a person's interruptibility in controlled environments and on small tasks. Psycho-physiological sensors have the advantage of providing more flexibility without being bound to a specific task, computer or location and are increasingly less invasive. Previous studies have shown that psycho-physiological features, such as electrodermal activity (EDA), heart rate or electroencephalographs (EEG), can be used to measure cognitive and emotional states (e.g., [9, 22]). Under the assumption that moments of high cognitive load or stress correlate with low interruptibility, studies have examined, for instance, the use of psycho-physiological sensors to calculate interruptibility from EEG data during a military exercise [36] or from a combination of heart rate variability and muscle activity on varying small tasks, such as word puzzles, mental arithmetic or racing games [12].

In our work, we aim to investigate the use of a combination of psycho-physiological sensors to automatically identify the interruptibility of a knowledge worker in a real-world working context. We build upon and extend the findings of previous research in the area by contributing two studies with software developers wearing psycho-physiological sensors: a lab study with participants working on real-world development tasks in a controlled environment, and a field study with participants working on their own tasks in their real-world office environments. In our analysis we focus on the use of psycho-physiological sensors to automatically infer the interruptibility of knowledge workers. In addition, we investigate the correlation between interruptibility, mental load and interruption lag and look at which tool support for interruptions developers desire.

The results of our studies provide evidence that psycho-physiological sensors can be used to classify the interruptibility of a software developer with high accuracy (91.5% for the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI 2015, April 18–23, 2015, Seoul, Republic of Korea.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3145-6/15/04 ...\$15.00.
<http://dx.doi.org/10.1145/2702123.2702593>

lab and 78.6% for the field study). Our results also show that we can build classifiers with high accuracy for a more fine-grained set of five states of interruptibility and that psycho-physiological data is very sensitive to individuals. In addition to these results, we provide evidence for the correlations between interruptibility, mental load and interruption lag and discuss potential tool support for software developers.

RELATED WORK

Interruptibility with Context-Aware Sensors

The greater part of related work that developed an interruptibility measure used context-aware sensors, such as audio and video streams, keyboard and mouse actions, active window information, table pressure, or task characteristics. For instance, Hudson et al. were able to classify interruptibility into two states (least interruptible vs. all other states) with 78% accuracy by simulating sensors and manually coding features from audio and video recordings, such as the number of people present, who was speaking, or whether the phone was on the hook [25]. Fogarty et al. also simulated sensors by manually encoding mouse and keyboard actions, such as highlighting a line or editing code. They measured interruptibility in terms of the interruption lag—the time between the notification and the start of an interruption—during coding tasks and achieved 72% accuracy for two state interruptibility classification (interruptible and engaged) [18]. Using a pressure sensor sheet on the desk, Tani et al. were able to achieve a similar two state interruptibility classification accuracy on typing and mouse operation with easy and hard phases [43]. Different to these, Iqbal et al. used task characteristics, such as the next subtask's difficulty, carry over of data across boundaries, and the percentage of parent task completion, to predict the cost of interruptions. They measured the cost based on resumption lag—the time needed to resume the primary task [27]. Ho et al. developed a context-aware mobile computing device to automatically detect activity transitions using accelerometers for measuring interruptibility. Their results show that messages delivered at activity transitions are better received than those delivered at random times [24].

Instead of context-aware sensors, our work uses biometric sensors to assess interruptibility. Especially with the recent advances and the development of low-cost biometric sensors with low-invasive form factors, a possible advantage is that these sensors are body-worn and not limited to laboratories, certain working environments or a specific software platform.

Biometric Sensors

An extensive amount of research investigates the link between psycho-physiological sensors and different cognitive and emotional states, such as high cognitive load or stress. The most studied sensors are electroencephalographs (EEG), eye tracking systems, sensors for electrocardiogram (ECG) or blood volume pulse (BVP), sensors measuring the electrodermal activity (EDA) and body temperature sensors.

EEG. EEG measures the aggregated electrical activity of the brain, which is caused when neurons fire. Different studies showed that certain frequency bands in the EEG data, called Alpha, Beta, Gamma, Delta and Theta can be linked to cognitive states, such as being focused, relaxed, or dreaming [9].

For instance, Gevins et al. found that an increase in theta activity and a decrease in alpha activity can be linked to an increase in memory load [19]. Kramer linked an increase in beta and decreases in alpha and theta to an increase in task engagement [30]. Several studies also used EEG devices to classify mental tasks or states of cognitive load [21, 32].

Eye Tracking. Eye trackers use the reflection of infrared light from the eyes to calculate the position of the visual focus and the pupil size. Interesting features are the pupil size, fixation duration or number of saccades. Particularly the pupil size (e.g. the peak amplitude of the pupil diameter) is an indicator for memory load or processing load, and varies with task difficulty [8]. Researchers showed that more difficult tasks demand longer processing time, induce higher subjective ratings of cognitive load and evoke greater pupillary response at salient subtasks [28]. Fixation durations and number of saccades are suitable to assess the designs of user interfaces [29].

ECG and BVP. For measuring the activity of the heart, either an ECG or BVP sensor can be used. ECG sensors measure the electrical activity of the heart using electrodes which are placed on the chest. BVP sensors, or photoplethysmographs, emit light which is absorbed by the oxyhemoglobin in the blood. The part of the light which is scattered back can be detected with a photodiode. ECG devices are more exact, but also more cumbersome to wear, therefore we chose a BVP sensor for our study. BVP can serve as a direct indicator for cognitive load [41], and can additionally be used to compute interbeat interval (IBI) and heart rate (HR).

EDA. EDA represents the skin conductivity that varies with sweating activity and can be measured by applying a small current with two electrodes. EDA has been linked to arousal, attention, emotional states, stress and anxiety [11]. In a study on text reading and arithmetic tasks imposing multiple cognitive load levels, a strong link between cognitive load and EDA was found [39].

Body temperature. Body temperature is influenced by emotions as well as stress. Vinkers et al. recently confirmed indications that stress influences body temperature in humans and found that body temperature rises with increasing stress [46]. In a study about autonomic nervous system response patterns evoked by emotions, skin temperature was demonstrated to be different in response to anger and fear [14].

Sensor combinations. Studies also applied multiple biometric sensors to measure cognitive load, task difficulty, task engagement and other cognitive states. For instance, Wilson analyzed mental workload in pilots during flight with multiple measures [47]. In prior work, we combined EDA, EEG and eye tracking to assess task difficulty in simple code comprehension tasks and found that a combination of all sensors to classify a new task achieved the highest accuracy of 84%. Haapalainen et al. combined an eye tracker, a heart rate monitor, ECG, EDA, EEG and body temperature sensors. They assessed the performance of different features to classify cognitive load on elementary cognitive tasks and found that the ECG median absolute deviation and median heat flux performed best, providing over 80% accuracy [22].

In our study, we use sensors for EEG, eye blinks, HR, BVP, EDA and body temperature due to the availability of low-cost and minimally invasive devices and their predictive power.

Interruptibility with Biometric Sensors

Fewer related work used psycho-physiological sensors to assess interruptibility. Kramer gathered EEG data during one hour of US military training and succeeded in classifying interruptibility based on labels gathered retrospectively [36]. Other researchers used measures of an eye tracker to compare mental workload during different hierarchic levels of task boundaries and were able to proof that mental workload dropped at high level task boundaries. This suggests that high level context switches are good moments for interruptions [5]. Chen et al. conducted a beeper study in which participants solved short tasks with varying difficulty. They measured heart rate variability as indicator for cognitive load and muscle activity through EMG and calculated interruptibility using linear regression [12]. They also developed a mobile phone, which classifies interruptibility into four states combining high / low mental load with high / low movement [13]. The studies conducted by Bailey et al. [5] and Chen et al. [12] were situated in a controlled laboratory environment and used well defined and relatively simple tasks, such as document editing based on specified comments, typing a given text or solving a word puzzle. Mathan et al [36] conducted a study during military training with more complex tasks, requiring different cognitive and also physical skills.

Our work differs in that it uses real-world software development tasks requiring a multitude of cognitive skills. In our field study, participants were wearing biometric sensors while working normally in their own offices on their own software development tasks with varying difficulty and context.

Interruption Management

Some related work already used interruptibility indicators in practical applications to avoid interruptions at unsuitable moments. For instance, a decision rule based software delivered unrelated instant messages only at times of context switches that were determined based on mouse movement and window switching. In their study they achieved a five times higher answer rate to messages compared to non-mediated message delivery [4]. Chen and Vertegaal automatically set the ring tone volume of a mobile phone using an interruptibility indicator based on heart variability and muscle activity. In a six person trial they found that participants were satisfied with the chosen notification level in 83% of the cases [12]. Only recently, novel designs to handle interruptions from incoming calls on smart phones have been developed adding new actions, such as ‘postpone’ or ‘send a message’, to traditional ones, such as ‘accept’ and ‘decline’ [10]. Our work can leverage existing interruption management support by providing an automated interruptibility measure using biometric sensors.

Interruption, Resumption and Edit Lag

There are three time spans commonly used in studies concerning the effects of interruptions: interruption, resumption and edit lag. Interruption lag is the timespan between a notification and the start of the interruption [3]. Resumption lag is

the timespan between the end of the interruption and the beginning of the suspended primary task, usually measured by monitoring the first mouse or keyboard action [1, 3]. Edit lag is the timespan between the end of the interruption and the first edit and represents a specialized measure of the resumption lag [40]. It is based on the assumption that interruptions in software development have a large effect and it takes minutes to regather context as opposed to resumption lag which is in the order of seconds [45].

For immediate interruptions (e.g. a phone call), the interruption lag is very short. Even short interruption lags (8s) are mitigating the disruptiveness of an interruption and can shorten the resumption lag [44]. For negotiated interruptions, the length of the interruption lag can be chosen by the interrupted person [37]. It can serve as indicator for interruptibility, following the notion of memory externalization before addressing an interruption [18]. The larger the memory load, the less interruptible a person is and the longer the interruption lag is required to find a suitable breakpoint. It has been shown that resumption lag is influenced by the availability of cues [3], the interruption length and demand [38], but not by stress, time pressure and flow [15].

In our work, we measured interruption lag, the traditional resumption lag and the specialized edit lag to report their lengths and analyze their correlation with interruptibility and mental load before the interruption.

STUDY DESIGN

To learn about the interruptibility of software developers and the use of biometric sensors to measure their interruptibility in a real-world context, we conducted two studies, a lab and a field study. The lab study was a first step to investigate whether biometric sensors can be used to measure interruptibility of software developers working on the same real-world change tasks in a controlled environment. As a second step, we conducted the field study to investigate how well results from the lab can be transferred to a real-world environment. In both studies, participants were wearing biometric sensors. Participants were interrupted at random times and asked to perform short arithmetic exercises, as well as rate their interruptibility, their mental load and the disturbance of the interruption. Figure 1 illustrates the study setup with one participant from the field study.

Psycho-Physiological Sensors

In both studies, we used two sensor devices: the Neurosky Mindband (<http://neurosky.com>), a headband to record electroencephalograph (EEG) and eye blink data, and the Empatica E3 wrist band (www.empatica.com) to record electrodermal activity (EDA), skin temperature, blood volume pulse, interbeat interval, and heart rate.

EEG and Eye-Related Measures

To measure the electrical brain activity and eye blinks we used the Neurosky Mindband, a low-cost and minimally invasive¹ headband with a one-channel EEG sensor. This headband works with one reference electrode at the ear and two

¹relative to other EEG sensor devices

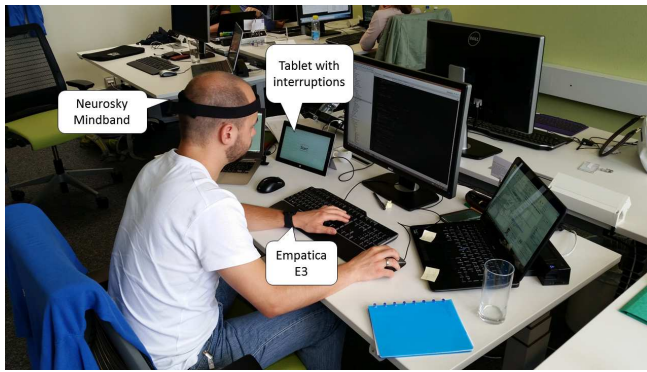


Figure 1. Study setup for one participant wearing the headband and wrist band in the field study. The tablet for triggering interruptions is placed next (left) to the participant’s main screen.

dry electrodes placed on the forehead reading signals mainly from the pre-frontal cortex. The device records the time-varying voltage signal sampled at 512Hz as a raw wave and as a wave filtered for noise. At the same time it records the signal quality that indicates the proper placement of the device. In addition, the headband also records two proprietary measures called Attention and Meditation that are both in the range from 0 to 100, sampled at 1Hz, and meant to indicate mental focus and mental calmness or relaxation respectively.

Skin- and Heart-Related Measures

To record skin- and heart-related measures, we used the wireless Empatica E3 wrist band that integrates an EDA sensor, a photoplethysmograph and a temperature sensor. The EDA sensor is used to measure skin conductance. It consists of two electrodes that are placed at the ventral area of the wrist. By applying a small current to the skin through these electrodes, skin conductance is measured down to $0.1\mu S$ at a frequency of 4Hz. The photoplethysmograph is an optical sensor for measuring blood volume pulse (BVP), which can be used to compute interbeat interval (IBI) and heart rate (HR).

Interruptions

To trigger interruptions during the studies we used a Windows Surface 2 tablet² that we placed close to the participant’s main monitor. For each interruption, the tablet played a sound and the display changed from a black to a white screen with a “Start” button on it. Participants were instructed to decide for themselves when to address the interruption—a technique for coordinating interruptions called “negotiated interruption” [37]. This technique usually works well and was used by others to simulate “normal” interruptions for software developers [18]. Using negotiated interruptions allows to measure interruption lag, which is the timespan between the notification and the moment a participant starts to address the interruption [2]. The interruption lag has previously been used as a measure for interruptibility, since researchers found that it corresponds to the time used to externalize the working memory before addressing the pending interruption [18].

Interruptions were composed of two parts, a mental arithmetic exercise and a set of questions on the participant’s perception of the interruption on five-point Likert scales. For the

arithmetic exercise participants were asked to solve a two-digit multiplication. As mental arithmetic exercises generally impose a high working memory load [33], they are considered an effective interruption for software developers [18]. After participants typed an answer into a text box on the tablet and clicked the “Ok” button, the correct solution was displayed to satisfy participants’ need for closure [31]. For the question set part, the tablet application prompted participants to rate their perceived disturbance from 1 (not at all disturbing) to 5 (very disturbing), their interruptibility at the time of the notification from 1 (highly interruptible) to 5 (not at all interruptible), and their mental workload at the time of the notification from 1 (very low) to 5 (very high). After answering these questions, the tablet application displayed a black screen and participants returned to their work.

Lab Study: Participants and Method

For the lab study, all participants worked on the same three real-world code change tasks in the same controlled environment. The study took place in a quiet office room with external distractions and interruptions, except for the ones triggered as part of the study, reduced to a minimum. All participants worked at the same computer with the same integrated development environment (IDE) setup for the study.

Through personal contacts, we recruited ten graduate students, one female and nine male. All participants had their major in computer science and an average of 4.1 years (± 3.8) professional and of 10.4 years (± 3.2) total development experience.

The study lasted 90 minutes per participant and had three parts, a preparation phase, a 60 minutes programming session and a brief follow-up interview. In the preparation phase, we asked each participant for some demographic information. Then we helped the participant to put on the two sensor devices and had the participant watch a movie of fish swimming in a fish tank for two minutes. The movie was intended to help participants relax and to record a baseline for each participant. We used this later on to normalize the captured data, for instance, to make heart rates comparable among participants with varying resting heart rates. To familiarize participants with the interruptions, we conducted a few test runs while displaying the fish tank movie. For the main part of the study, each participant was asked to work for 60 minutes in the Eclipse IDE on three code change tasks that we explained to them beforehand. During this programming session participants were frequently interrupted after random time intervals that were between one and eleven minutes long. These time intervals model interruption frequency occurring in reality [20]. At the end of the study, we briefly interviewed each participant. In the interview, we asked participants about their perception of the primary tasks, the interruptions with the peripheral arithmetic tasks, as well as more generally about the disruptiveness of interruptions and tool support they would desire for better managing interruptions.

Project and Tasks

We chose the drawing framework JHotDraw³ for the three code change tasks. JHotDraw is an open source project that

²<http://www.microsoft.com/surface/en-us/products/surface-2>

³<http://www.jhotdraw.org/>

has evolved over several years, is well structured and allows for easy testing due to its graphical user interface. The three study tasks were chosen to represent real code change tasks with varying difficulty levels to stress various levels of mental load and various states of interruptibility in a participant.

Adding Circles. The first task is to add a drawable figure, namely a circle, to the Draw application. The task requires to add a button with a provided icon to the toolbar and code to draw the circle. As there is already a feature for drawing an ellipse, code can be reused and the main difficulty is to identify the right places where new code needs to be inserted.

Adding Hexagons. The second task is similar to the first one, but requires to add a hexagon instead of a circle. Knowledge obtained in the first task could be reused, however, drawing a hexagon is more difficult and requires knowledge in geometry. As an optional help, a document with explanations of the geometry of a hexagon was provided.

Adding Text and URL. The third task is to add text and a clickable URL into an existing message dialog. The main task difficulty is to locate the right place in the code for implementing the functionality, and to get familiar with the Java API on message dialogs as well as user interface components.

To validate that the perceived task difficulty varied between tasks, we asked participants to rate them from 1 (very easy) to 5 (very difficult). They rated the first and third task as rather easy (2.4 ± 0.7 and 2 ± 1.4) and the second as rather difficult (3.9 ± 1.0).

Field Study: Participants and Method

To learn more about interruptibility and the use of psycho-physiological sensors in the field, we conducted a study with ten professional software developers working on their own tasks and in their real-world office environments. For this study, we visited professional developers in their work places and studied them for two hours each. We did not restrict any external influences, such as interruptions by co-workers or distractions, and we did not limit the work or the participant's work setup, such as the activities they performed during work, the IDEs they used or the office layout.

We recruited ten professional software developers (1 female, 9 male) between their early twenties and late forties from four software development companies of varying size. Participants were recruited through personal contacts and recruiting emails. They had an average of 8.5 years (± 7.5) professional software development experience and an average of 12.7 years (± 6.0) of total development experience.

The field study lasted 2.5 hours per participant and had again three parts, a preparation phase, a main study session of 2 hours and a brief follow-up interview. As in the lab study, we first asked each participant for some demographic information. We then helped them to put on the sensors and had them watch the fish tank movie to help them relax, record a baseline and familiarize participants with the interruptions. For the main study session, participants worked on their usual tasks while being frequently interrupted after random time intervals that, again, were between one and eleven minutes long.

For this session participants were told to work as usual during their work day without restriction on their activities, such as checking emails or browsing the web, and to switch tasks as they would normally do. Also, we told all participants and co-workers beforehand to interact during the session as they would usually do during their work day. At the end of the study, we again conducted a brief follow-up interview with participants on the perception of the tasks, the disruptiveness of interruptions and desired tool support.

Tasks

During the main study session, participants worked on a variety of tasks, such as the elimination of a performance bottle neck in a web application, the implementation of a user interface component, or the implementation of test cases for a business application. Most participants worked mainly on one primary task during the two hour session and only rarely switched to other small tasks. For these tasks, they used a variety of tools, such as IDEs (e.g. Visual Studio or Eclipse), revision control tools (e.g. SourceTree), web browsers (e.g. Firefox) and email or calendar clients (e.g. MS Outlook).

Data Collection and Analysis

During the course of both studies we collected data from the psycho-physiological sensor devices and data on participants' computer interaction captured through monitoring and observation. In addition, we collected participants' answers to the questions during the interruptions, their demographic information and notes from the brief follow-up interviews. We recorded a total of 30 hours of psycho-physiological data, 10 hours for the 10 lab study participants and 20 hours for the 10 field study participants. The psycho-physiological data for two lab study participants was not recorded successfully for the entire session. Therefore, we will only present the analysis and results of the 72 valid interruption samples collected from the 8 other lab study participant. The distribution of interruptions is fairly constant per study, with an average of 9 (± 1.6) per person for the 1h lab study and 13.9 (± 2.7) for the 2h field study.

Psycho-Physiological Data

For both studies, we captured the psycho-physiological data with the same sensors and applied the same data cleaning and analysis steps. We discuss these in the following.

EEG and Eye. The raw signal from the EEG sensor is sampled at 512 Hz. We applied a 50 Hz notch filter to remove noise and then split the signal into five commonly used brain wave frequency bands using Matlab's `pwelch` function: α (8-12Hz), β (12-30Hz), γ (30-80Hz), δ (0-4Hz), and θ (4-8Hz) [23]. Additionally, we computed fractions of all combinations of frequency bands and the two ratios $\theta/(\alpha + \beta)$ and $\beta/(\alpha + \theta)$ that have previously been shown to carry information on a person's mental activity [30, 32]. Following a technique suggested by Manilov [35], we extracted eye blinks from the raw signal using a Butterworth filter and a peak finding algorithm and calculated the number of eye blinks per time unit. Finally, we used the pre-computed Attention and Meditation signals from Neurosky and extracted the minimum, maximum, mean and standard deviation.

Skin and Heart. The EDA signal generally serves as a measure for arousal and has two main components, the low frequency tonic part that changes over a period of minutes and the higher frequency phasic part that changes within seconds [42]. After filtering noise by applying an exponential smoothing filter, we used a Butterworth filter to split the EDA signal into its tonic and phasic part. From the tonic signal we extracted the skin conductance level (SCL). From the phasic signal we extracted features related to the peaks, in particular the number of peaks per time unit, the mean and the sum of peak amplitudes, and also calculated the area under the curve (AUC). Based on the skin temperature data captured from the integrated thermometer, we extracted the mean temperature and amplitude features, such as the mean and maximum. Based on the captured BVP data we calculated several features, such as the number of peaks per time unit, the mean peak amplitude as well as the heart rate, its mean and variance. Finally, from the IBI, we computed features of heart rate variability, such as the standard deviation of the signal (SDNN) and the percentage of successive IBIs with a difference greater than 20ms (PNN20) and 50ms (PNN50) [34]. All extracted features have previously been linked to various cognitive and emotional states (see Related Work).

Normalization. Since we train our machine learning classifier across participants and psycho-physiological data is very individual, we normalized the data per participant. Therefore, we collected baseline measures during the time each participant watched the relaxing fish tank movie. Normalizing a feature's value for an interruption was then done by subtracting the feature's value calculated using the baseline data from the one calculated using the interruption data.

Interruption, Edit and Resumption Lag

To calculate interruption, edit and resumption lags for participants, we needed to capture five time stamps for each interruption: when the notification for the interruption occurred ($T_{notification}$), when the participant started to address the interruption ($T_{IntStart}$), when the participant finished with the interruption (T_{IntEnd}), when the participant continued to work on her or his primary task using a mouse click or a keyboard action ($T_{FirstInteract}$) and when the participant made the first edit after the interruption ($T_{FirstEdit}$). Interruption lag can then be calculated as $T_{IntStart} - T_{notification}$, resumption lag as $T_{FirstInteract} - T_{IntEnd}$ and edit lag as $T_{FirstEdit} - T_{IntEnd}$. We captured the notification and the interruption start and end with the tablet application. For the lab study, we captured the first interaction and edit with a monitoring software that we installed on the lab study computer and which recorded screen shots and logged mouse and keyboard actions. For the field study we captured the two time stamps by observing participants interactions.

Outcome Measures

To classify interruptibility, both the interruptibility ratings and the interruption lag could serve as ground truth. Although the interruption lag might seem to be a more objective measure, we decided to use the rating in the classification for two reasons. First, the majority of prior work used the same rating and we wanted to provide a comparable measure, and second,

even though interruption lags and interruptibility ratings correlate, they are distributed differently. While ratings are distributed binomially, interruption lags are distributed exponentially. This supports our observation during the studies with some participants addressing notifications quickly, regardless of their interruptibility. Therefore we believe that the ratings represent the interruptibility more accurately and independent of the participant's interruption handling behavior.

Since we are also interested in examining whether mental load and interruptibility are positively correlated as previous work suggests [7] and whether interruptions during high mental load are more disturbing, we collected participants' ratings of mental load and perceived disturbance as secondary measures. In addition, we collected the interruption, edit and resumption lags as described above to examine findings of prior work on their relation to interruptibility.

RESULTS

In this section we report on the results of automatically measuring interruptibility using psycho-physiological sensors, the links between interruptibility, mental load and the various lags as well as on the timing and support for interruptions desired by participants.

Measuring Interruptibility

To investigate the use of psycho-physiological sensors for measuring the interruptibility of knowledge workers in a real-world context, we applied machine learning to the collected data in a post-hoc analysis. In particular, we are interested in classifying interruptibility into two states (interruptible or not) as well as a more fine-grained classification using five interruptibility states from highly interruptible to not at all interruptible. For the two state classification we categorized data ratings from 1 (highly interruptible) to 5 (not at all interruptible) into two groups by labeling data with ratings of 1 to 3 as interruptible and ratings 4 and 5 as not interruptible. This categorization results in software developers being interruptible in 75% of the samples (83% for the lab and 71% for the field study). For the five state analysis we labeled data with participants' five point ratings of interruptibility. We analyzed the two data sets collected from the lab study ($n=72$ interruption samples) and the field study ($n=139$) separately.

Time Window and Classification Algorithm

To examine which time window of psycho-physiological data per interruption works best for the classification, we applied machine learning to several time windows, ranging from ten seconds to three minutes all ending with the notification (see Figure 2). Taking into consideration two and five state classification as well as lab and field study results, a time window of ten seconds works generally better than longer ones, except for the five states case of the field study in which longer time windows perform better, probably due to more frequent noise artifacts. The results also show that there are no big differences in accuracy across various time windows. In addition, we examined the use of three different machine learning algorithms: Naïve Bayes, Decision Trees and Support Vector Machine based on Weka's implementations. Overall, Naïve Bayes outperformed the other two algorithms, which is why we will focus on Naïve Bayes in the following.

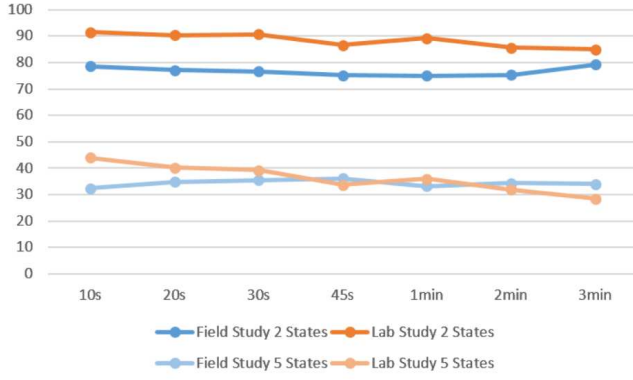


Figure 2. Classification accuracies for two and five categories using different time windows and Naïve Bayes.

Validation Methods

For the classification we used a Naïve Bayes implementation of the Weka machine learning framework [48]. We applied ten times ten-fold cross-validation, where instances of the participants were distributed randomly across the folds using stratification (‘per instance’ cross-validation). We chose this approach to investigate the feasibility of using psycho-physiological sensors to predict interruptibility for a development team. This method could be used in a real-world scenario by initially gathering about two hours of data per developer on the team to train a classifier and then using it to classify interruptibility based on real-time sensor data for developers on the team. To gather further insights on the generalizability and the sensitivity of the sensors across individuals, we also performed ‘per participant’ cross-validation. To prevent overfitting, we applied a nested correlation-based feature selection technique (Weka’s CfsSubsetEval) that chooses features with high correlation with the class variable but low correlations among each other, using data from the training set of each fold.

Two and Five States Interruptibility

The results of our approach using Naïve Bayes and a ten second time window of psycho-physiological data for each interruption are presented in Table 1. For both validation methods, we calculated the accuracy, Cohen’s Kappa and the standard deviation of the accuracies obtained during each fold and run. For both studies, we are able to classify the interruptibility at the sample points with high accuracy into two states (91.5% for the lab and 78.6% for the field study) as well as into five states (43.9% for the lab and 32.5%) for ‘per instance’ cross-validation. These classifiers perform significantly better than simple majority classifiers for both studies and number of states, except for five states classification in the field. We believe the lack of significance in the latter case is due to more frequent noise artifacts in the field that make it more difficult to distinguish between the finer grained levels. On the other hand, ‘per participant’ cross-validation does not result in a significant performance difference compared to a majority classifier and reveals a large variance among participants.

Tables 2 and 3 show the confusion matrices for the classification per instance, along with F-measures depicting individual class prediction performance. White cells indicate correct

Truth	Prediction	
	interruptible	not interruptible
interruptible	58.9 / 88.2	1.1 / 10.8
not interruptible	5 / 19	7 / 21
F_{lab}	0.95	0.70
F_{field}	0.86	0.58

Table 2. Confusion matrix for Naïve Bayes classification into two states using per instance cross-validation for lab and field study (lab / field) with individual class accuracies (F-measure)

Truth	Prediction				
	1	2	3	4	5
1	1.2 / 2.4	1.9 / 12.4	2.8 / 2.2	1.1 / 1	0 / 1
2	1.5 / 3.2	9.5 / 22	13.4 / 6.5	0.5 / 2.3	0.1 / 11
3	1.9 / 2.9	7.7 / 12.7	18 / 5	0.4 / 3.1	0 / 11.3
4	1 / 1.9	2 / 4.3	1.2 / 6.6	0.2 / 5.3	2.6 / 6.9
5	0.1 / 0.1	0.1 / 2.3	1 / 1.7	1.1 / 0.5	2.7 / 10.4
F_{lab}	0.19	0.41	0.56	0.04	0.52
F_{field}	0.16	0.45	0.18	0.28	0.37

Table 3. Confusion matrix for Naïve Bayes classification into five states using per instance cross-validation for lab and field study (lab / field) with individual class accuracies (F-measure)

predictions, colored cells indicate wrong predictions, where darker cells correspond to more severe errors. For the classification into five states, the confusion matrix reveals that a great part of the errors is caused by prediction of an adjacent state, and only few errors are severe, which is very promising.

Table 4 presents the features that were selected at least once by the nested feature selection and for which classifier they were used for in the per instance classification. The table shows that β , γ and the mean temperature were valuable in all scenarios. EDA related features were only chosen for the lab study data, which might be due to the more frequent occurrence of noise artifacts in the field and the short time windows chosen. Overall, these results also show that all sensors can provide value for predicting interruptibility.

Interruptibility, Mental Load and Lags

In both studies, there were significant and high positive correlations between the participants’ ratings of the perceived disturbance, the mental load and the interruptibility (Pearson’s $r > 0.7$, $p < 0.0001$). For instance, participants’ ratings on interruptibility were highly correlated with ratings on mental load in the lab ($r=0.815$) and in the field ($r=0.702$), as well as with ratings on disruptiveness in the lab ($r=0.807$) and in the field ($r=0.741$). These high correlations support the often assumed link between mental load and interruptibility, that moments of low mental load are suitable for interruptions, and that interruptions during moments of low mental load are perceived less disturbing [7].

Interruptibility was also positively correlated with the interruption lag for the lab study (Pearson’s $r_{lab}=0.382$, $p < 0.001$) and the field study ($r_{field}=0.282$, $p < 0.001$). Participants generally took advantage of the so-called negotiated interruption, with an average interruption lag of 30.4 seconds (± 7.5) in the lab study, and 44.9 seconds (± 6.7) in the field study. Furthermore, 17 of 20 participants commented that they use the interruption lag to finish the current edit in most cases, which

# States	Study	Per Instance Cross-Validation			Per Participant Cross-Validation			Majority Classifier
		Accuracy	Cohen's Kappa	Stdev	Accuracy	Cohen's Kappa	Stdev	Accuracy
Two	Lab	91.5%*	0.65	0.7	74.9%	-0.11	36.0	83.3%
	Field	78.6%*	0.44	1.3	69.4%	0.22	19.7	70.7%
Five	Lab	43.9%*	0.18	2.9	37.6%	0.11	21.7	38.9%
	Field	32.5%	0.13	2.1	28.2%	0.07	14.2	32.4%

Table 1. Classification results by number of states and study (* indicates that there is a significant difference in accuracy to the majority classifier).

Sensor	Used Features
EEG	α [2L, 5L], α/γ [2F], α/δ [2F], β [2L, 2F, 5L, 5F], β/α [2L, 2F, 5L], β/γ [2F, 5F], β/δ [2F, 5F], β/θ [2L, 5L], γ [2L, 2F, 5L, 5F], γ/α [2L, 2F, 5L], γ/β [2F, 5F], γ/δ [2F, 5F], γ/θ [2F, 5L], δ [5L], δ/β [2L, 5L], δ/γ [2L, 2F, 5L], δ/θ [2L, 5F], θ [2L], θ/α [2L], θ/δ [2L], $\beta/(\alpha + \theta)$ [2F, 5F], Mean Attention [2F, 5F], Stdev Attention [2F], Min Meditation [2L]
Photoplethysmograph	Mean Peak Amplitude BVP [2F], Sum Peak Amplitude BVP [2L, 5L], Max Peak Amplitude BVP [5F], Mean HR [2F], IBI PNN20 [2L, 2F], IBI NN50 [2F],
Temperature	Mean Temperature [2L, 2F, 5L, 5F]
EDA	Mean Phasic Peak Amplitude EDA [2L, 5L], Sum Phasic Peak Amplitude EDA [2L, 5L], Phasic Peak Frequency EDA [2L, 5L]

Table 4. Most predictive features for Naïve Bayes classification for per instance cross-validation, and their use in the classifiers (2L/2F: lab/field study two states, 5L/5F: lab/field study five states).

overlaps with Fogarty et al.'s finding that participants externalize their working memory before addressing a negotiated interruption [18]. These findings—the positive correlation between interruptibility and interruption lag and the evidence for a longer interruption lag corresponding to a higher working memory load—further support the link between interruptibility and working memory load and thus mental load.

Researchers also found a possible link between resumption/edit lag and interruption lag or interruptibility [44]. We did not find any strong support for this link across studies. There was only a significant correlation between resumption and interruption lag ($r=0.275$, $p=0.001$) in the field study and a significant correlation between edit and interruption lag ($r=0.251$, $p=0.04$) in the lab study. No other significant correlations at the level of 0.05 (two-sided) were found, including correlations with participants' ratings of interruptibility, mental load or disruptiveness.

Interruption Timing and Support

We used the follow-up interviews to learn more about the cost of interruptions at certain moments and possible tool support. In particular, we asked participants to rate certain situations that we identified in previous literature from 1 (strongly like) to 5 (strongly dislike) and found that, similar to findings of previous studies, participants like interruptions at the end of a task (1.5 ± 0.6) but not in the middle (4 ± 0.7), as well as they dislike them when the mental load is high (5 ± 0) and/or the current task is difficult (4.4 ± 0.9). In situations where participants are stuck and are not making any progress, they feel more mixed about interruptions (2.9 ± 1.2) and several partic-

ipants mentioned to dislike interruptions in these situations although they stated that interruptions would usually be beneficial for their task and for gaining a different perspective.

When we asked participants about the kind of support they desired for interruptions, they mentioned a tool that displays your interruptibility to co-workers, for instance, by using a lamp (mentioned by 7 participants), and a tool that turns interruptions on or off based on the current mental load (mentioned by 5 participants). In particular, support for in-person interruptions is more needed than for computer based ones since they are generally perceived more disruptive and cannot be ignored. However, participants also commonly mentioned that important interruptions should not be blocked even in situations of extremely high mental load, and that the company culture should be respected by the tool, e.g., a tool should not prevent interactions that foster team spirit.

DISCUSSION

The primary focus of the presented work was to investigate the use of psycho-physiological sensors to measure the interruptibility of knowledge workers in a real-world context. The results from our lab and our field study show that using these sensors, we are able to generate machine learning classifiers that can identify a software developer's state of interruptibility—for two as well as five states—with high accuracy. The fact that the results for the lab study are better than the ones for the field study is possibly an indicator for the effect that external influences can have on such sensors, but might also stem from the different sample sizes. The poor performance of 'per participant' cross-validation indicates the high sensitivity of psycho-physiological sensors to individual differences. We assume that much more data is needed to investigate whether it is possible to generalize interruptibility classification using psycho-physiological sensors across different individuals. We believe that our primary choice of 'per instance' cross-validation, for which we train a classifier with data from one team, represents a reasonable trade-off between effort, limitations and value and is applicable in a real-world scenario.

The overall high accuracy for both studies, the task variety and real-world office environment in the field study as well as the use of representative real-world tasks in the lab study, show that these sensors have great potential for measuring interruptibility in a real-world context. As main usage we imagine to display the interruptibility state in real-time via a "traffic light" lamp or IM status, which can potentially help avoiding costly personal interruptions at inopportune moments. Another possibility is to automatically adapt notification settings based on the current interruptibility, where the priority of interruptions has to be taken into account to not miss im-

portant ones. Especially the small time windows of ten seconds required to measure interruptibility with a high accuracy show that these sensors provide the possibility of a real time interruptibility index that can be used for such purposes.

Although our studies were limited to the software development domain, the mobility of these sensors allows for their use in a broad range of domains without being bound to a specific task, computer platform or location and are technically less restrictive than more context-aware sensors, such as a table top sensor or computer interaction monitors (e.g., [18, 43]). We believe that our results are therefore an encouraging step in the use of such sensors in a real-world work context and warrant further research on applying these sensors to a broader range of knowledge workers.

The results show that measures from the EEG, Photoplethysmograph, temperature and EDA sensors can all provide valuable information for classifying interruptibility. While EEG measures were selected for all classifiers, the EEG headband might be too obtrusive for long-term use. In future work, we plan to examine the use of subsets of sensors over longer periods of time to achieve a high accuracy and usability.

In addition to providing evidence for the benefits of using these sensors to measure interruptibility, the results of our study also confirm previous results on the correlation between mental load and interruptibility [7, 26], the link between interruption lag and interruptibility in setups with negotiated interruptions [18], and the findings that interruptions at task-boundaries are less disruptive [5].

CONCLUSION

In this work we investigated the use of psycho-physiological sensors to automatically classify the interruptibility of knowledge workers in a real-world context. We conducted two studies, a lab and a field study, in which we captured the psycho-physiological data of twenty participants for a total of 30 hours and interrupted them at random times. Using a Naïve Bayes classifier, we are able to predict the interruptibility of participants with high accuracy, improving significantly upon a majority classifier. Our results also confirm previous findings on the positive correlation between interruptibility and mental load, which further supports the use of psycho-physiological sensors that have already been shown to indicate states of mental load in other studies.

For future work, we aim at designing and prototyping tool support leveraging the predictive power of psycho-physiological sensors to help knowledge workers with their management of interruptions. Based on the interviews of our study, such support is highly desired for direct interruptions from co-workers, but also for computer-based interruptions, such as e-mails and instant messages.

ACKNOWLEDGMENTS

The authors would like to thank all study participants. This work was funded in part by SNF and an ABB research grant.

REFERENCES

- Adamczyk, P. D., and Bailey, B. P. If not now, when?: the effects of interruption at different moments within task execution. In *Proc. CHI'04* (2004).
- Altmann, E. M., and Trafton, J. G. Memory for goals: An activation-based model. *Cognitive science* 26, 1 (2002).
- Altmann, E. M., and Trafton, J. G. Task interruption: Resumption lag and the role of cues. Tech. rep., DTIC Document, 2004.
- Arroyo, E., and Selker, T. Attention and intention goals can mediate disruption in human-computer interaction. In *Proc. INTERACT'11*. 2011, 454–470.
- Bailey, B. P., and Iqbal, S. T. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM TOCHI* 14, 4 (2008).
- Bailey, B. P., and Konstan, J. A. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior* 22, 4 (2006).
- Bailey, B. P., Konstan, J. A., and Carlis, J. V. The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In *Proc. INTERACT'01* (2001), 593–601.
- Beatty, J. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin* 91, 2 (1982).
- Berger, H. Über das elektroencephalogramm des menschen. *Archiv für Psychiatrie und Nervenkrankheiten* 87, 1 (1929), 527–570.
- Böhmer, M., Lander, C., Gehring, S., Brumby, D. P., and Krüger, A. Interrupted by a phone call: exploring designs for lowering the impact of call notifications for smartphone users. In *Proc. CHI'14*, ACM (2014).
- Boucsein, W. *Electrodermal activity*. Springer, 2012.
- Chen, D., Hart, J., and Vertegaal, R. Towards a physiological model of user interruptibility. In *INTERACT'07*. Springer, 2007.
- Chen, D., and Vertegaal, R. Using mental load for managing interruptions in physiologically attentive user interfaces. In *Ext. Abstract CHI'04*, ACM (2004).
- Collet, C., Vernet-Maury, E., Delhomme, G., and Dittmar, A. Autonomic nervous system response patterns specificity to basic emotions. *Journal of the autonomic nervous system* 62, 1-2 (1997).
- Conard, M. A., and Marsh, R. Single and multiple interruptions increase task performance time, but don't affect stress, pressure or flow. 2010.
- Czerwinski, M., Horvitz, E., and Wilhite, S. A diary study of task switching and interruptions. In *Proc. CHI'04* (2004).
- Fogarty, J., Hudson, S. E., Atkeson, C. G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J. C., and Yang, J. Predicting human interruptibility with sensors. *ACM TOCHI* 12, 1 (2005).

18. Fogarty, J., Ko, A. J., Aung, H. H., Golden, E., Tang, K. P., and Hudson, S. E. Examining task engagement in sensor-based statistical models of human interruptibility. In *Proc. CHI'05*, ACM (2005).
19. Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., and Rush, G. Monitoring working memory load during computer-based tasks with eeg pattern recognition methods. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 40, 1 (1998).
20. González, V. M., and Mark, G. Constant, constant, multi-tasking craziness: managing multiple working spheres. In *Proc. CHI'04*, ACM (2004).
21. Grimes, D., Tan, D. S., Hudson, S. E., Shenoy, P., and Rao, R. P. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In *Proc. CHI'08*, ACM (2008).
22. Haapalainen, E., Kim, S., Forlizzi, J. F., and Dey, A. K. Psycho-physiological measures for assessing cognitive load. In *Proc. Ubicomp*, ACM (2010).
23. Handy, T. C. *Event-related potentials: a methods handbook*. MIT Press, 2005.
24. Ho, J., and Intille, S. S. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *Proc. CHI'05*, ACM (2005).
25. Hudson, S., Fogarty, J., Atkeson, C., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J., and Yang, J. Predicting human interruptibility with sensors: a wizard of oz feasibility study. In *Proc. CHI'03*, ACM (2003).
26. Iqbal, S. T., and Bailey, B. P. Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. In *Ext. Abstract CHI'05*, ACM (2005).
27. Iqbal, S. T., and Bailey, B. P. Leveraging characteristics of task structure to predict the cost of interruption. In *Proc. CHI'06*, ACM (2006).
28. Iqbal, S. T., Zheng, X. S., and Bailey, B. P. Task-evoked pupillary response to mental workload in human-computer interaction. In *Ext. Abstracts CHI'04*, ACM (2004).
29. Jacob, R. J., and Karn, K. S. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind* 2, 3 (2003), 4.
30. Kramer, A. F. Physiological metrics of mental workload: A review of recent progress. *Multiple-task Performance* (1991).
31. Kruglanski, A. W. Lay epistemic theory in social-cognitive psychology. *Psychological Inquiry* 1, 3 (1990).
32. Lee, J. C., and Tan, D. S. Using a low-cost electroencephalograph for task classification in hci research. In *Proc. UIST'06*, ACM (2006).
33. Lemaire, P. The role of working memory resources in simple cognitive arithmetic. *European Journal of Cognitive Psychology* 8, 1 (1996).
34. Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., and Schwartz, P. J. Heart rate variability standards of measurement, physiological interpretation, and clinical use. *European heart journal* 17, 3 (1996).
35. Manoilov, P. Eye-blinking artefacts analysis. In *Proc. Compsystech* (2007), 52.
36. Mathan, S., Whitlow, S., Dorneich, M., Ververs, P., and Davis, G. Neurophysiological Estimation of Interruptibility: Demonstrating Feasibility in a Field Context. In *Proc. ICACS'07* (2007).
37. McFarlane, D. Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction* 17, 1 (2002).
38. Monk, C. A., Trafton, J. G., and Boehm-Davis, D. A. The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology: Applied* 14, 4 (2008), 299.
39. Nourbakhsh, N., Wang, Y., Chen, F., and Calvo, R. A. Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In *Proc. OzCHI'12*, ACM (2012).
40. Parnin, C., and Rugaber, S. Resumption strategies for interrupted programming tasks. *Software Quality Journal* 19, 1 (2011).
41. Peper, E., Harvey, R., Lin, I.-M., Tylova, H., and Moss, D. Is there more to blood volume pulse than heart rate variability, respiratory sinus arrhythmia, and cardiorespiratory synchrony? *Biofeedback* 35, 2 (2007).
42. Schmidt, S., and Walach, H. Electrodermal activity (eda): State-of-the-art measurement and techniques for parapsychological purposes. *Journal of Parapsychology* 63, 3 (1999).
43. Tani, T., and Yamada, S. Estimating user interruptibility by measuring table-top pressure. In *Ext. Abstracts CHI'13*, ACM (2013).
44. Trafton, J. G., Altmann, E. M., Brock, D. P., and Mintz, F. E. Preparing to resume an interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies* 58, 5 (2003).
45. van Solingen, R., Berghout, E., and van Latum, F. Interrupts: just a minute never is. *IEEE software* 15, 5 (1998).
46. Vinkers, C. H., Penning, R., Hellhammer, J., Verster, J. C., Klaessens, J. H., Olivier, B., and Kalkman, C. J. The effect of stress on core and peripheral body temperature in humans. *Stress* 16, 5 (2013).
47. Wilson, G. F. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology* 12, 1 (2002).
48. Witten, I. H., and Frank, E. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.